Sudharshan S Vazhkudai

email (preferred): vazhkuda@mcs.anl.gov
url: http://www.mcs.anl.gov/~vazhkuda

Math and Computer Science
Argonne National Laboratory
9700 South Cass Avenue
Argonne, IL 60439
Office: 630-252-3397 (until 12/2002)

RESEARCH INTERESTS

- Wide-Area Resource Management systems, Computational/Data Grids, Distributed Data-Intensive Science
- Volunteer Computing, Distributed Operating Systems
- Scheduling, Load Sharing Environments, Co-Allocations, Performance Predictions, Replication
- Incentives for resource management, Reservations, Contracts/Guarantees, Information services/dissemination

RESEARCH SUMMARY

My interests are primarily in the area of Distributed Resource Management. To this end, I am currently involved in Globus/GriPhyN Data Grid development - built techniques to locate data in highly replicated environments; developed performance prediction strategies for bulk Data Grid transfers aiding data transfer schedules; building co-allocation middleware enabling parallel and fast access to replicated data, etc. Prior to this, I led the design and development of a performance oriented Linux Cluster (with kernel level modifications) to achieve enhanced throughput.

TEACHING INTERESTS

Develop and teach courses in grid computing, Internet scale distributed systems, distributed operating Systems, and foundation courses in computer science. More precisely, construct and teach a series of courses in distributed systems, progressing from tightly coupled distributed OSs to clusters to massively distributed wide-area systems, volunteer computing systems and runaway virtual organizations.

EDUCATION

- Ph.D. candidate (CS; expected May 2003), University of Mississippi/Argonne National Laboratory
  Dissertation: *Bulk Data Transfer Forecasts and the Implications to Grid Scheduling*
  Highlight: *Three years of Dissertation fellowship from Argonne National Laboratory (US Department of Energy Lab); Thesis work on the R&D award winning Globus wide-area distributed systems toolkit; Reviewed papers for IEEE & ACM Conferences, IEEE TPDS.*
  Committee: *Dr. Tobin Maginnis (Chair), Dr. Ian Foster (ANL/U of Chicago), Dr. Conrad Cunningham and Dr. Dawn Wilkins*

- Master of Science (CS) University of Mississippi, 1998
  Thesis: *Performance Oriented Distributed OS-Evolutionary Steps towards a Distributed Linux*
  Highlight: *Led a team of five graduate students in the design & development of a Distributed Linux; Inducted into into ACM's Upsilon Pi Epsilon Honor Society.*
  Advisor: *Dr. Tobin Maginnis*

- Bachelor of Engineering (CS), Karnatak University, India, 1996
  Project: *Yash - Yet Another Shell, An Enhanced C-Shell for the Unix Environment*
  Highlight: *Ranked among top 5% of the class.*

AWARDS

- Ph.D. Dissertation Fellowship (2001 - 2002) - **Argonne National Laboratory** (*Research in Data Grids*)

- Wallace Givens Fellowship (Summer 2000) - **ANL/US Department of Energy** (*Awarded each year to 3 select candidates from a group of approximately 40 students*)
- Inducted into Upsilon Pi Epsilon and Gamma Beta Phi Honor societies (*GPA based*)

WORK EXPERIENCE

- ***Research Associate*, Math and Computer Science, Argonne National Laboratory, Argonne, IL**
  Argonne National Laboratory is a U.S. Department of Energy research center specializing in Grid Computing. The Globus project, being developed at ANL, is evolving as the de facto standard in Grid Computing, winning the R&D 100 award and being widely used in major production systems (European Data Grid, Earth Systems Grid, Particle Physics Data Grid, etc.).
  [January 2001 - December 2002]
  Working on resource management issues in the Globus Data Grid (disparate, massively distributed systems):
  - Developed Replica selection strategies in widely replicated environments
  - Built performance prediction models of data transfer times using extensive statistical analysis
  - Built strategies for the efficient discovery of information
  - Building and analyzing techniques for co-allocating transfers (parallel downloads, etc.)
  - *Further information available in the "Research Work" section; Relevant papers in "Publications" section*

- ***Wallace Givens Research Fellow*, Math and Computer Science, Argonne National Laboratory, Argonne, IL**
  [May 2000 - September 2000]
  - Developed a prototype resource selection broker, using ClassAds, for the Globus Data Grid
  - Built performance instrumentation of the GridFTP data movement tool
  - Developed information service providers for GridFTP performance data discovery
  - This work resulted in a publication in CCGRID 2001

- ***Senior Technical Consultant*** *(concurrent appointment)*, **Sair Linux GNU Certification, Oxford, MS**
  Sair Linux GNU Certification specializes in administering Linux Systems candidacy exams world wide.
  [May 1999 - June 2001]
  - Designed and developed certification exams (along with a core team) for testing practical Linux systems experience
  - Developed training manuals, tests, lab exercises in advanced Linux system, network and security administration

- ***Summer Intern*, Environmental Systems Research Institute (ESRI Inc.), Redlands, CA**
  ESRI was founded in 1969 as a research group devoted to improving methods of handling geographically referenced data.
  [May 1998 - August1998]
  - "Purified" (staunching leaks) ARC/INFO, a leading Geographic Information Systems (GIS) product comprising of legacy code written in FORTAN, C and C++
  - Performed bug fixing for ARC/INFO
  - Ported ARC/INFO libraries to various architectures

- ***Research Assistant*, Department of Computer Science, University of Mississippi, Oxford, MS**
  National Classroom Project, supported by Microsoft Corporation, is committed to bringing computing into classrooms.
  [August 1997 - December 1997]
  - Worked on the development of an in-class portable lab that provides learning enhancement services to students and assessment services to teachers
  - Developed a Web based dynamic testing/teaching Environment for C++, WWW Remote Compilation environment, etc., for portable hand-held devices
  - Developed numerous libraries for the portable C++ compiler ("eC") for hand-held devices

- ***Design Engineer*, Research and Development, Wipro Infotech Limited, India**
  Wipro Infotech is a leading IT Company in India working on diverse software issues.
  [August 1996 - August 1997]
  - Designed and developed Windows Printer Drivers for Epson based printers along with a core development team

- Developed several driver functions invoked by the GDI (device specifics, modes, copy data, etc.), color calibration schemes and intensity and brightness control
- Modified Unidrv driver development kit to include Epson specific commands and utilities, with support from firmware team

TEACHING EXPERIENCE

- ***Instructor*, Department of Computer Science, University of Mississippi**
  [August 1998 - May 2000]
  Structure and Methodology:
  - Prepared and delivered class lectures
  - Developed homework assignments (up to 10), quizzes (weekly), programming projects (up to 5), course projects (spanning the semester), and exams (3)

  Courses Taught:
  - Survey of Computing (CSCI 103; Fall 99 - Spring 2000)
  - Computer Organization (CSCI 223; Summer 99)
  - Programming for Scientists - FORTRAN (CSCI 251; Fall 98 - Spring 99)

  Further information on teaching experience and interests available in the "Teaching Statement"

SELECTED RESEARCH WORK

- **Ph.D. Dissertation** (*exploring resource management issues in the Globus/GriPhyN Data Grids*)
  Description:
  - Prototyped storage resource management systems and associated storage resource selection brokers
  - Defined a set of services for efficient resource selection
  - Developed techniques for predictive analysis, resource co-allocation, etc.
  - Validation and comparison using various Globus Data Grid applications (Particle Physics Data Grid and the Earth Systems Grid)

  Insights:
  - An understanding of Grid community requirements in the context of distributed data intensive science
  - Exposure to several state-of-the-art Grid related software systems (such as: Globus, NWS, ClassAds, GridFTP, MDS, LDAP, etc.)
  - Exposure to techniques and tradeoffs in building and analyzing disparate grid systems
  - An understanding of statistical tools in the context of systems performance analysis, predictions (regression analysis), etc.

  Deliverables:
  (As "technology transfer", several of these are part of the Globus Toolkit Release; Research papers available from Globus Web site)
  - Prototype Brokers for Globus Data Grids
  - Performance Instrumentation and discovery service for the GridFTP bulk data movement tool
  - Univariate math predictors, multivariate regression-based predictors
  - Static and dynamic load balancing of co-allocated data flows for parallel downloads

- **Performance Oriented Linux Cluster**
  Description:
  - Led the design and development of a Distributed Linux
  - Built a High-speed Communication Protocol (Communication Manager) for a Cluster of Linux Machines involving multiple channels and short-circuiting protocol stack
  - Built a Distributed File System exploiting Linux VFS layers and prefetching
  - Designed and guided the development of Global IPC mechanism and group communication
  - Developed a remote process execution environment

  Insights:
  - In-depth understanding of Distributed OS concepts
  - Code-level understanding of Linux (i)Datalink Layer (Ethernet Driver), (ii)VFS Layer, (iii)Process Management & Scheduling
  - Understanding of intricacies in Kernel Programming and Debugging.

  Deliverables:
  - Research papers and Distributed Linux Patches for 2.0 kernel

- **Framework for Distributed Protocols**
  Description:
  - Explored the construction of a framework for distributed decision making protocols (leader election, distributed scheduling, etc.)
  - Applied design patterns such as layered architecture, delegation and templates in constructing a white-box framework
  - Developed an initial abstraction of vertical and horizontal views comprising of initiators/listeners and application, protocol-specific and platform-dependent layers respectively

  Insights:
  - Applying design patterns and frameworks to software development, hot-spot analysis

  Deliverables:
  - Research papers and an initial prototype framework in C++

PUBLICATIONS

- **Journal**

1. S. Vazhkudai, J. Schopf, "Using Regression Techniques to Predict Large Data Transfers", *Submitted to the Journal of High Performance Computing Applications - Special Issue on Grid Computing: Infrastructure and Applications.*

2. S. Vazhkudai, J.M. Syed, P.T. Maginnis, "PODOS - The Design and Implementation of a Performance Oriented Linux Cluster", *Journal of Future Generation Computer Systems - Special Issue on Cluster Computing*, Volume 18, Issue 3, pp. 335-352, January 2002.

- **Conferences/Workshops**

1. S. Vazhkudai, J. Schopf, "Using Disk Throughput data in Predictions of End-to-End Grid Transfers", *To appear in 3rd International Workshop on Grid Computing (GRID 2002),* Baltimore, Maryland, November 2002.

2. S. Vazhkudai, J. Schopf, "Predicting Sporadic Grid Data Transfers", *Proceedings of the 11th IEEE International Symposium on High Performance Distributed Computing (HPDC-11),* pp. 188-196, Edinburgh, Scotland, July 2002.

3. S. Vazhkudai, J. Schopf, I. Foster, "Predicting the Performance of Wide-Area Data Transfers", *Proceedings of the 16th International Parallel and Distributed Processing Symposium (IPDPS 2002),* Fort Lauderdale, Florida, April 2002.

4. S. Vazhkudai, S. Tuecke, I. Foster, "Replica Selection in the Globus Data Grid", *Proceedings of the IEEE International Conference on Cluster Computing and the Grid (CCGRID 2001),* pp. 106-113, Brisbane, Australia, May 2001.

5. R. Buyya, S. Vazhkudai, "Compute Power Market: Towards a Market-Oriented Grid", *Proceedings of the IEEE Session on Global Computing on Personal Devices,* pp. 574-581, Brisbane, Australia, May 2001.

6. S. Vazhkudai, G.V. Laszewski, "A Greedy Grid - The Grid Economic Engine Directive", *Proceedings of the IEEE Workshop on Internet Computing and E-Commerce,* San Francisco, California, April 2001.

7. S. Vazhkudai, S. Tuecke "A Storage Broker for the Globus Environment - A ClassAd Based Implementation", *Poster in Supercomputing 2000* , Dallas, Texas, Nov 2000.

8. S. Vazhkudai, P.T. Maginnis, "The PODOS File System - Exploiting the High-Speed Communication Subsystem", *Proceedings of the IEEE International Workshop on Cluster Computing Technologies, Environments, & Applications*, pp. 453-459, Las Vegas, Nevada, June 2000.

9. S. Vazhkudai, H.C. Cunningham, "A Reusable Software Framework for Distributed Decision-Making

Protocols", *Proceedings of the Workshop on Distributed Objects in Computational Science (DOCS'2000)* , pp. 867-873, Las Vegas, Nevada, June 2000.

10. S. Vazhkudai, P.T. Maginnis, "A High Performance Communication Subsystem for PODOS", *Proceedings of the First IEEE International Conference on Cluster Computing*, pp. 81-91, Melbourne, Australia, December 1999.

- **Technical Reports**
  1. S. Vazhkudai, "Bulk Data Transfer Forecasts and the Implications to Grid Scheduling", *Ph.D. Dissertation*, University of Mississippi/Argonne National Laboratory. (*In Preparation*)

  2. S. Vazhkudai, "Distributed Linux: Evolutionary Steps", *Masters Thesis, Technical Report CISE TR 1998-22*, University of Mississippi, December 1998.

- **Book Chapters**
  1. Contributed chapters to *Linux System Administration, Linux Networking & Linux Security and Ethics* by Tobin Maginnis, Published by John Wiley & Sons

- **Papers In Preparation**
  1. S. Vazhkudai, "Enabling the Co-Allocation of Grid Data Transfers"

TALKS

1. **University of Southern California**, California: "Server Selection in Data Grids", April 2001.
2. **Fermilabs**, Illinois: "A Matchmaking Approach to Replica Selection", January 2001. ***(Invited)***
3. **Atlanta Linux Showcase**, Georgia: "The Design and Evolution of Communication in PODOS", October 1999.
4. **LinuxWorld Expo Conference**, California: "Transmission Group based Communication for PODOS", August 1999.

GRANT WRITING

- *Performance Oriented Distributed Linux* - (PI: Dr. Tobin Maginnis) Largely responsible for developing and writing an NSF proposal for the construction of a distributed Linux testbed, exploring concepts in cluster communication, distributed file systems, load sharing environments and fabric support of large scale systems.

PROFESSIONAL and LEADERSHIP ACTIVITIES

- Reviewed papers for: *High Performance Distributed Computing (HPDC-11), IEEE Transactions for Parallel and Distributed Systems, 3rd International Workshop on Java for Parallel and Distributed Computing, IEEE International Symposium on Cluster Computing and the Grid (CCGRID 20001), 4th International Conference on Algorithms and Architectures for Parallel Computing, IEEE International Conference on Cluster Computing (CLUSTER 2000), International Workshop on Grid Computing (Grid 2000), European Cluster Computing Conference (EUROPAR 2000), Parallel and Distributed Systems Track of ACM SAC 2000*
- Reviewed books ***"Kernel Projects for Linux", "Linux System Administration"*** , Published by Addison Wesley for correctness, feasibility of projects, providing suggestions and ideas for further exercises and topics.
- Member of **ACM**, **IEEE**, **Task Force on Cluster Computing (TFCC), Global Grid Forum**
- Led a team of five graduate students in the construction of a Distributed Linux
- Led a team of three senior-level students in the construction of a Distributed Shell
- Initiated a research project on the exploration of patterns and frameworks in the context of distributed protocol development

SOFTWARE SYSTEMS DEVELOPED (*Educational Purposes*)

- **Distributed Systems:** PODOS (*Distributed Linux*).
- **Operating System Components:** Yash (*Enhanced C Shell for the UNIX environment*).

- **Networks:** Remote Folder Sharing Environment (*perform basic directory operations, compilation, etc., on remote machines*).
- **Compilers:** Shell Script To C Converter (*Grammar for a Scripting language and a Converter*).
- **Databases:** Query Processor for Relational Algebra expressions.
- **WWW:** Web based Testing Environment, CGI Library Kit.
- **Component Architectures:** Chaos (*IDE for C projects under Unix using Implicit Invocation of objects*), File Manager for UNIX (*Windows like*).

PERSONAL DETAILS

- Citizenship: India
- Visa Status: F1 - Student Visa

REFERENCES

1. **Dr. Tobin Maginnis**, Associate Professor, Department of Computer and Information Science, University of Mississippi, 237 Kinard Hall, University, MS 38677, (662) 801-0459, ptm@pix.cs.olemiss.edu

2. **Dr. Gregor von Laszewski**, Assistant Computer Scientist, Mathematics and Computer Science Division, Argonne National Laboratory, and Fellow of Computation Institute, University of Chicago, 9700 S. Cass Ave., Bldg. 221, Argonne, IL 60439, (630) 252-0472, gregor@mcs.anl.gov

3. **Dr. H. Conrad Cunningham**, Associate Professor and Chairperson, Department of Computer and Information Science, University of Mississippi, 237 Kinard Hall, University, MS 38677, (662) 915-5358, cunningham@cs.olemiss.edu

4. **Dr. Robert P. Cook**, Yamacraw Professor, Department of Mathematics and Computer Science, Georgia Southern University, MPP Building #3008, 201 Georgia Ave, Statesboro, GA 30460, (912) 681-0892, bobcook@gasou.edu

5. **Dr. Dawn Wilkins**, Associate Professor, Department of Computer and Information Science, University of Mississippi, 237 Kinard Hall, University, MS 38677, (662) 915-7309, dwilkins@cs.olemiss.edu